

## Exercise: Banking-Hive Queries

Amith Kumar  
1NT18IS027  
'A'

## GitHub:

[https://github.com/Amith027/BigData\\_lab\\_2021/tree/main/Basic%20Hive%20Queries%20Banking](https://github.com/Amith027/BigData_lab_2021/tree/main/Basic%20Hive%20Queries%20Banking)

# CSV file

LibreOffice

File Edit View Insert Format Styles Sheet Data Tools Window Help

banking.csv

Liberation Sans 10 B I U T A

A1 fx X ✓ tid

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	tid	bname	cname	amount										
2	506	yes	amith	45000										
3	507	canara	kumar	17000										
4	508	icici	nikhil	32000										
5	509	union	aftab	20000										
6	510	vanija	bharath	10990										
7														
8														
9														
10														
11														
12														
13														
14														
15														
16														
17														
18														
19														
20														
21														
22														
23														
24														
25														
26														
27														
28														
29														
30														
31														

Sheet 1 of 1

Default English (UK) Insert 100%

Average: Sum: 0

## Queries:

1. Create database called customer.
2. Create a table called bank(tid, bname, cname, amount).
3. Insert the values using INSERT statement.

```
hive> create database customer;
OK
Time taken: 0.126 seconds
hive> use customer;
OK
Time taken: 0.029 seconds
hive> create table bank(tid int, bname string, cname string, amount float) row format delimited fields terminated by ",";
OK
Time taken: 0.199 seconds
hive> desc bank;
OK
tid                int
bname              string
cname              string
amount             float
Time taken: 0.06 seconds, Fetched: 4 row(s)
hive> insert into bank values(501,"hdfc","hari",89000),(502,"vijaya","krishna",50000),(503,"canara","shashi",46000),(504,"corporation","prasanna",25000),(505,"union","manka",55000);
Query ID = hdoop_20210710160124_1de3b57d-eff2-4d67-98bf-98b31c36ebf6
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1625904496337_0009, Tracking URL = http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8088/proxy/application_1625904496337_0009/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625904496337_0009
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-10 16:01:36,103 Stage-1 map = 0%, reduce = 0%
2021-07-10 16:01:45,567 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.16 sec
2021-07-10 16:01:52,810 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.55 sec
MapReduce Total cumulative CPU time: 7 seconds 550 msec
Ended Job = job_1625904496337_0009
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://127.0.0.1:9000/user/hive/warehouse/customer.db/bank/.hive-staging_hive_2021-07-10_16-01-25_014_6686694486055925908-1/-ext-10000
Loading data to table customer.bank
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.55 sec HDFS Read: 19618 HDFS Write: 545 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 550 msec
OK
Time taken: 30.486 seconds
hive> select * from bank;
OK
501    hdfc    hari    89000.0
502    vijaya  krishna 50000.0
503    canara  shashi  46000.0
504    corporation  prasanna 25000.0
505    union    manka    55000.0
Time taken: 0.275 seconds, Fetched: 5 row(s)
```

## 4. Import the values using CSV file.

```
hive> load data local inpath '/home/jayesh/Desktop/banking.csv' into table bank;
Loading data to table customer.bank
OK
Time taken: 1.07 seconds
hive> select * from bank;
OK
501    hdfc    hari    89000.0
502    vijaya  krishna 50000.0
503    canara  shashi  46000.0
504    corporation  prasanna 25000.0
505    union    manka    55000.0
NULL    bname    cname    NULL
506    yes     amith    45000.0
507    canara  kumar    17000.0
508    icici   nikhil   32000.0
509    union    aftab    20000.0
510    vanijya  bhath    10990.0
Time taken: 0.277 seconds, Fetched: 11 row(s)
hive>
```

## 5. Query operations like groupby, aggregate, orderby and having clause, view.

```
hive> select tid,bname,cname,amount from bank where cname="amith";
OK
506   yes    amith    45000.0
Time taken: 0.235 seconds, Fetched: 1 row(s)
hive> select cname,amount from bank where amount>=30000.00 group by cname,amount;
Query ID = hdoop_20210710171447_9b42c536-5df4-4dfa-b215-90b482c99916
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1625904496337_0010, Tracking URL = http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8088/proxy/application_1625904496337_0010/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1625904496337_0010
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-10 17:15:04,425 Stage-1 map = 0%,   reduce = 0%
2021-07-10 17:15:23,329 Stage-1 map = 100%,   reduce = 0%, Cumulative CPU 4.3 sec
2021-07-10 17:15:36,789 Stage-1 map = 100%,   reduce = 100%, Cumulative CPU 8.85 sec
MapReduce Total cumulative CPU time: 8 seconds 850 msec
Ended Job = job_1625904496337_0010
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Reduce: 1   Cumulative CPU: 8.85 sec   HDFS Read: 13137 HDFS Write: 246 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 850 msec
OK
amith    45000.0
hari     89000.0
krishna  50000.0
manka    55000.0
nikhil   32000.0
shashi   46000.0
Time taken: 51.861 seconds, Fetched: 6 row(s)
hive> █
```

```
>
> select tid,cname,sum(amount) from bank group by tid,cname having sum(amount)>40000;
Query ID = hdoop_20210710172346_bcfe795e-995b-4a8d-a85f-19b8fe402480
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1625904496337_0013, Tracking URL = http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8088/proxy/application_1625904496337_0013/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1625904496337_0013
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-10 17:23:57,691 Stage-1 map = 0%,   reduce = 0%
2021-07-10 17:24:03,964 Stage-1 map = 100%,   reduce = 0%, Cumulative CPU 3.34 sec
2021-07-10 17:24:12,261 Stage-1 map = 100%,   reduce = 100%, Cumulative CPU 9.15 sec
MapReduce Total cumulative CPU time: 9 seconds 150 msec
Ended Job = job_1625904496337_0013
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Reduce: 1   Cumulative CPU: 9.15 sec   HDFS Read: 14829 HDFS Write: 239 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 150 msec
OK
501     hari    89000.0
502     krishna 50000.0
503     shashi  46000.0
505     manka   55000.0
506     amith   45000.0
Time taken: 27.679 seconds, Fetched: 5 row(s)
hive> █
```

```

Time taken: 53.64 seconds, Fetched: 1 row(s)
hive> select max(amount) as MaxAmt, min(amount) as MinAmt, avg(amount) as AvgAmt from bank;
Query ID = hdoop_20210710171755_9987f311-036b-4cb5-82c2-45d7b203244f
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1625904496337_0011, Tracking URL = http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8088/proxy/application_1625904496337_0011/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1625904496337_0011
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-10 17:18:12,450 Stage-1 map = 0%, reduce = 0%
2021-07-10 17:18:23,142 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.1 sec
2021-07-10 17:18:46,855 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.91 sec
MapReduce Total cumulative CPU time: 8 seconds 910 msec
Ended Job = job_1625904496337_0011
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.91 sec HDFS Read: 17129 HDFS Write: 123 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 910 msec
OK
89000.0 10990.0 38999.0
Time taken: 53.64 seconds, Fetched: 1 row(s)
hive>
>
>
> select tid,cname,amount from bank where amount>25000 order by amount DESC;
Query ID = hdoop_20210710172006_6541c45f-7fef-4676-ab1c-26078c790f12
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1625904496337_0012, Tracking URL = http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8088/proxy/application_1625904496337_0012/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1625904496337_0012
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-10 17:20:20,766 Stage-1 map = 0%, reduce = 0%
2021-07-10 17:20:28,402 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.26 sec
2021-07-10 17:20:34,671 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 9.29 sec
MapReduce Total cumulative CPU time: 9 seconds 290 msec
Ended Job = job_1625904496337_0012
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 9.29 sec HDFS Read: 12567 HDFS Write: 270 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 290 msec
OK
501 hari 89000.0
505 manka 55000.0
502 krishna 50000.0
503 shashi 46000.0
506 amith 45000.0
508 nikhil 32000.0
Time taken: 30.333 seconds, Fetched: 6 row(s)
hive>

```

```

hive> create view cust_transaction as select tid,cname from bank;
OK
Time taken: 0.29 seconds
hive> select * from cust_transaction;
OK
501 hari
502 krishna
503 shashi
504 prasanna
505 manka
NULL cname
506 amith
507 kumar
508 nikhil
509 aftab
510 bharath
Time taken: 0.226 seconds, Fetched: 11 row(s)
hive>

```

## 6. Create one more table

```
hive> create table customer(cust_id int, tid int, location string) row format delimited fields terminated by ",";
OK
Time taken: 0.188 seconds
hive> desc customer;
OK
cust_id      int
tid          int
location     string
Time taken: 0.115 seconds, Fetched: 3 row(s)
hive> insert into customer values(101,501,"Bengaluru"),(102,503,"Mysore"),(103,502,"Hubli"),(104,505,"Dharwad");
Query ID = hdoop_20210710172947_b0938cfb-5c3a-4c03-919d-75ffcd86a16f
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1625904496337_0014, Tracking URL = http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8088/proxy/application_1625904496337_0014/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625904496337_0014
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-10 17:29:59,030 Stage-1 map = 0%, reduce = 0%
2021-07-10 17:30:06,344 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.5 sec
2021-07-10 17:30:13,628 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.91 sec
MapReduce Total cumulative CPU time: 8 seconds 910 msec
Ended Job = job_1625904496337_0014
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://127.0.0.1:9000/user/hive/warehouse/customer.db/customer/.hive-staging_hive_2021-07-10_17-29-47_435_2089794662186451562-1/-ext-10000
Loading data to table customer.customer
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.91 sec HDFS Read: 16977 HDFS Write: 399 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 910 msec
OK
Time taken: 27.843 seconds
hive> select * from customer;
OK
101      501      Bengaluru
102      503      Mysore
103      502      Hubli
104      505      Dharwad
Time taken: 0.236 seconds, Fetched: 4 row(s)
hive> █
```

## 7. Join Operation

```
hive> select cname,location from bank b, customer c where b.tid=c.tid;
Query ID = hdoop_20210710173356_1149fab8-c86b-4e0e-9d49-28a985225c43
Total jobs = 1

SLF4J: Found binding in [jar:file:/home/hadoop/hadoop-3.2.1/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2021-07-10 17:34:09 Starting to launch local task to process map join; maximum memory = 239075328
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1625904496337_0015, Tracking URL = http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8088/proxy/application_1625904496337_0015/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625904496337_0015
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-07-10 17:34:25,620 Stage-3 map = 0%, reduce = 0%
2021-07-10 17:34:33,954 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 5.21 sec
MapReduce Total cumulative CPU time: 5 seconds 210 msec
Ended Job = job_1625904496337_0015
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 5.21 sec HDFS Read: 9648 HDFS Write: 192 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 210 msec
OK
hari      Bengaluru
krishna   Hubli
shashi    Mysore
manka     Dharwad
Time taken: 39.993 seconds, Fetched: 4 row(s)
hive> █
```



## 8. Alter Operation

```
hive> alter table customer rename to cus;
OK
Time taken: 0.237 seconds
hive> desc cus;
OK
cust_id          int
tid              int
location         string
Time taken: 0.053 seconds, Fetched: 3 row(s)
hive> alter table cus add columns(phno int);
OK
Time taken: 0.16 seconds
hive> insert into cus values(207,103,"shivmoga",9876543211);
Query ID = hdoop_20210710174124_73af9ea8-5f26-4cf7-a3e2-b894b2e7295b
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1625904496337_0016, Tracking URL = http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8088/proxy/application_1625904496337_0016/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1625904496337_0016
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-10 17:41:36,652 Stage-1 map = 0%, reduce = 0%
2021-07-10 17:41:45,008 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.88 sec
2021-07-10 17:41:52,277 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 9.36 sec
MapReduce Total cumulative CPU time: 9 seconds 360 msec
Ended Job = job_1625904496337_0016
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://127.0.0.1:9000/user/hive/warehouse/customer.db/cus/.hive-staging_hive_2021-07-10_17-41-24_911_6312734894819831375-1/-ext-10000
Loading data to table customer.cus
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 9.36 sec HDFS Read: 18112 HDFS Write: 346 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 360 msec
OK
Time taken: 30.015 seconds
hive> select * from cus;
OK
101  501  Bengaluru  NULL
102  503  Mysore    NULL
103  502  Hubli    NULL
104  505  Dharwad  NULL
207  103  shivmoga 1286608619
Time taken: 0.207 seconds, Fetched: 5 row(s)
hive> █
```